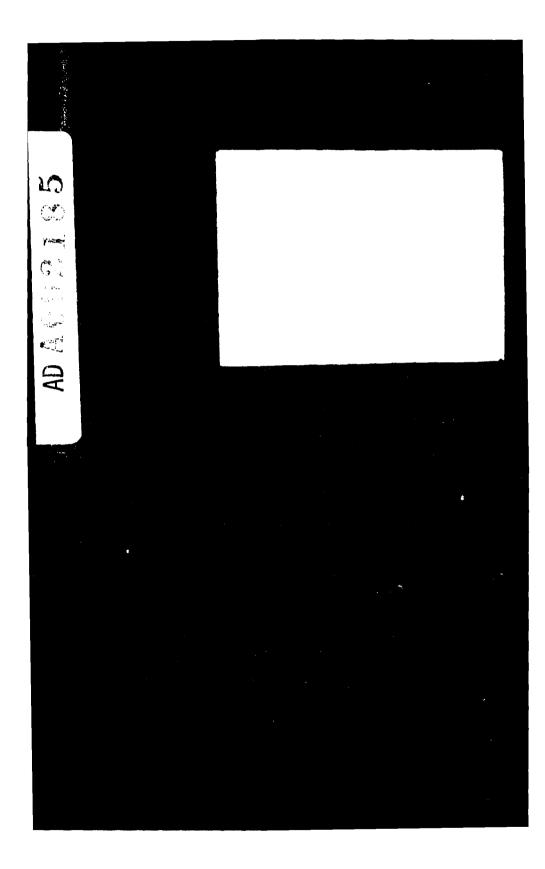MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

ON SCHUR OPTIMALITY

by

Gregory M. Constantine
Department of Mathematics
Indiana University
Bloomington, Indiana 47405

March, 1980

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR TR- 80 - 1117	2. GOVT ACCESSION NO. AD-A090185	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ON SCHUR OPTIMALITY.		5. TYPE OF REPORT & PERIOD COVERED Interim rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Gregory M. Constantine		8. CONTRACT OR GRANT NUMBER(s) AFOSR-76-3050
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Indiana Department of Mathematics Bloomington, Indiana 47405		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE March 1980
		13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Schur-optimality, Schur-convexity, D-, A-, E-optimality criteria,
Linear model, Incomplete block designs, majoritation, eigenvalues.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Schur-optimality is defined (in the general setting of a linear model) as a
generalization of the well-known D-, A- and E-optimality criteria. Techniques
to establish Schur-optimality are outlined, based chiefly on a process of
averaging information matrices and on vector majorization. A design with a
completely symmetric information matrix of maximal trace is shown to be
Schur-optimal. A design with an information matrix of maximal trace and
exactly two distinct nonzero eigenvalues is proved Schur-better than a large
class of designs. One description of a subcollection of designs over which

Schur-optimality holds is given only in terms of the diagonal elements of the information matrices. Consequences of this are than examined in the setting on block designs.

ON SCHUR OPTIMALITY

by

Gregory M. Constantine
Department of Mathematics
Indiana University
Bloomington, Indiana 47405

(In amintirea lui Grigore)

SUMMARY

Schur-optimality is defined (in the general setting of a linear model) as a generalization of the well-known D-, A- and E-optimality criteria. Techniques to establish Schur-optimality are outlined, based chiefly on a process of averaging information matrices and on vector majorization. A design with a completely symmetric information matrix of maximal trace is shown to be Schur-optimal. A design with an information matrix of maximal trace and exactly two distinct nonzero eigenvalues is proved Schur-better than a large class of designs. One description of a subcollection of designs over which Schur-optimality holds is given only in terms of the diagonal elements of the information matrices. Consequences of this are then examined in the setting of block designs.

---

1. INTRODUCTION

Oftentimes, in experiments, interest arises in estimating
parameters with equal (or near equal) precision. A design
balanced as much as possible for the parameters of interest
is intuitively felt to be the right choice since no para-
meter ought to be left to disadvantage. But precision (for-
mulated in terms of the covariance matrix of the estimates)
turns out not to be related to the concept of balance nei-
ther in a direct way nor in an obvious one. While balance
relates directly to the entries of the information matrix,
precision is closely connected to its spectrum. In the
present paper a connection between balance and precision is
made with the help of the concept of Schur-optimality.
Certain convex operations on the entries of the information
matrix (as trends for balance) turn out to be compatible
with the Schur-convex functions defined on the spectrum
of the information matrix, as measures of precision. By
compatibility we roughly mean that the more balanced a
design tends to be, the closer to a minimum the values of
the Schur-convex functions get. Our goal is to eventually find
a design for which these Schur-convex functions reach their
minima. As we define it, the concept of Schur-optimality is
very strong and a design can only seldom be proved Schur-
optimal over the collection of all possible designs. It is
however quite convenient to show that whenever a design with

many desirable symmetries exists, it is better (in a very strong sense) than a large number of less symmetric designs. This is the main scope of the paper.

A design d describes the way a certain stastical experiment is to be conducted. It generally specifies where the observations are to be taken and in what proportion at each location. The design is, within limits, subject to choice by the experimenter.

Let Ω be the collection of all possible designs. For $d \in \Omega$ we assume the following expected linear response:

$$E(Y) = X_d \theta$$

where Y is the $m \times 1$ vector of uncorrelated observations with common variance σ^2, X_d is the design matrix of dimension $n \times p$ and θ is a $p \times 1$ vector of unknown parameters. Usually statistical interest arises in estimating linear functions of θ_1, a $v \times 1$ subvector of θ. Then the reduced normal equations for θ_1 can be written as

$$C_d \theta_1 = Q_d' Y$$

with Q_d a $m \times v$ matrix and C_d a $v \times v$ nonnegative definite matrix, called the information matrix of the design d (for θ_1). Unless otherwise specified, all the information matrices C_d in the sequel will be with reference to the subvector θ_1.

In the block design setting, for example, we are to compare v varieties (labeled $1, 2, \ldots, v$) via b blocks

of size k ($< v$). A design d in this case is a $k \times b$ array with varieties as entries and blocks as columns. The collection of all designs is denoted by $\Omega_{v,b,k}$. The usual additive model, under which these designs are considered, specifies the expectation on variety i in block j as $\alpha_i + \beta_j$ where α_i is the (unknown) effect of variety i and β_j is the (unknown) effect of the jth block. Let

$$\theta = (\alpha_1, \ldots, \alpha_v, \beta_1, \ldots, \beta_b)'$$

be the vector of unknown parameters. We are interested only in the subvector $\alpha = (\alpha_1, \ldots, \alpha_v)'$ of variety effects. The information matrix for α, when the design d is used for estimation, is

$$C_d = diag(r_{d1}, \ldots, r_{dv}) - \frac{1}{k} N_d N_d'$$

where r_{di} is the number of replications of variety i in d and $N = (n_{dij})$, with n_{dij} signifying the number of times variety i occurs in block j. The above information matrix is nonnegative definite with row sums zero, for all $d \in \Omega_{v,b,k}$. The row sums being zero reflects the fact that only linear contrasts (i.e. functions $\ell'\alpha$ with $\ell'\underset{\sim}{1} = 0$, where $\underset{\sim}{1} = (1, \ldots, 1)'$) are estimable under any design d. This is very often the case in discrete settings. The collection of information matrices C_d, with $d \in \Omega_{v,b,k}$, has therefore a common kernel generated by $\underset{\sim}{1}$.

Denote by $\mu(C_d)$ the nondecreasingly ordered vector of eigenvalues of C_d outside an eventual common kernel (which

is of no help in distinguishing among designs). Let $\mu(C_d)$
be a vector of length $n(\leq v)$. In most relevant instances
n is either v or $v-1$. Designs d for which $\mu(C_d)$
has some zero entries are ruled out as bad designs, generally
because they are disconnected. We shall therefore focus our
attention to designs d for which $\mu(C_d)$ has all its entries
positive. In summary, to a design $d \in \Omega$ we associate an in-
formation matrix C_d and the nondecreasingly ordered vector
$\mu(C_d)$ of its eigenvalues associated with the eigenvectors
of C_d outside a kernel common to all C_d with $d \in \Omega$, i.e.,

$$\mu(C_d) = (\mu_{d1}, \ldots, \mu_{dn})' \quad ; \quad 0 < \mu_{d1} \leq \cdots \leq \mu_{dn} .$$

Statistical optimality criteria (e.g. D-,A-,E-criteria)
were defined on the collection of vectors $\mu(C_d)$, with $d \in \Omega$
by Ehrenfeld (1955) and Kiefer (1959). A design $d*$ is called

D-,A- or E-optimal if it minimizes $\prod_{i=1}^{v-1} \mu_{di}^{-1}$, $\sum_{i=1}^{v-1} \mu_{di}^{-1}$ or
μ_{d1}^{-1} respectively, over all $d \in \Omega$. Minimization of functions
of the form $\sum_{i=1}^{v-1} f(\mu_{di})$ with f convex has been considered
by Kiefer (1975) and Cheng (1978). While D- and A-criteria
can be easily expressed in this latter form, the E-criterion
cannot be expressed in such form. It can only be recovered
as a limit of such functions. In the first part of this paper
we extend criteria of the form $\sum_{i=1}^{v-1} f(\mu_{di})$ with f convex
to $\psi(\mu(C_d))$ with ψ Schur-convex and nonincreasing in its ar-
guments. The three functions associated with D-,A- and E-cri-

teria are all genuine instances of Schur-convex functions which
are nonincreasing in their arguments. We then define the concept
of a Schur-optimal design and, in section 3, outline general
methods of establishing Schur-optimality. The techniques which
we mention rely chiefly on convexity (averaging information
matrices) and on vector majorization. In the last section
of the paper we illustrate the material presented in section 3
by showing that a design with a completely symmetric matrix of
maximal trace is Schur-optimal. This result very much resembles
Proposition 1 of Kiefer (1975) on Universal optimality. There
is no direct relationship between Schur-optimality and Universal
optimality but the concepts are discussed comparatively in sec-
tion 4. We then show that a design with an information matrix
of maximal trace and exactly two distinct nonzero eigenvalues
is Schur-optimal over classes of designs which satisfy a
verifiable condition on the entries of their information matrices.
For a subclass of designs a sufficient condition is conve-
niently formulated only in terms of the diagonal entries of
the information matrices. In the block design setting corol-
laries are derived for binary designs with exactly two nonzero
eigenvalues. Examples of such designs are the Partially Bal-
anced Incomplete Block designs with two associate classes, as
well as extended and abridged both Balanced Incomplete Block
designs and Group Divisible designs.

2. DEFINITIONS

Let us recall the following:

A square matrix is called doubly stochastic if it has nonnegative entries with row and column sums equal to 1.

Let I be an interval on the real line R. A function ψ defined on I^n with real values is called Schur-convex if

$$\psi(Sx) \leq \psi(x)$$

for all $x \in I^n$ and all S doubly stochastic.

A real function ψ defined on I^n is said to be nonincreasing in its arguments if it is a nonincreasing function when restricted to each of its arguments.

A function F defined on I^n is called symmetric if $F(Px) = F(x)$ for all $x \in I^n$ and all permutation matrices P.

A function ϕ defined on a convex set A in R^n with real values is called convex if

$$\phi(\alpha x + (1-\alpha)y) \leq \alpha\phi(x) + (1-\alpha)\phi(y)$$

for all $x, y \in A$ and all $0 \leq \alpha \leq 1$.

By observing that a permutation matrix (and its inverse) is doubly stochastic one can see that a Schur-convex function is always symmetric. The converse is of course false, but the well-known result of Birkhoff and Von Neumann which states that the collection of doubly stochastic matrices is the convex span of permutation matrices provides us with many examples of Schur-convex functions. By this result it immediately follows that any symmetric and convex function is Schur-convex. Lastly, a Schur-convex function need not be convex, e.g.,

$$\psi(x_1, x_2) = |x_1 - x_2|^{1/2}.$$

Let b be a constant such that $\mu_{d1} \leq b$ for all $1 \leq 1 \leq n$ and all $d \in \Omega$, e.g., b can be the maximal trace of C_d with $d \in \Omega$. The constant b is always finite, either because Ω is finite or because of arguments involving compactness. Let $I = [0, b]$. Then $\mu(C_d) \in I^n$ for all $d \subset \Omega$. So we can define $\psi(\mu(C_d))$ for ψ Schur-convex.

For convenience we set by definition $\psi(C_d)$ be $\psi(\mu(C_d))$ for ψ Schur-convex and nonincreasing in its arguments. We are now ready to define Schur-optimality.

Definition 2.1. A design d^* is said to be Schur-better than another design d (notation $d^* \succeq d$) if

$$\psi(C_{d^*}) \leq \psi(C_d)$$

for all Schur-convex functions ψ nonincreasing in their arguments.

Definition 2.2. A design d^* in Ω is called Schur-optimal over Ω if $\psi(C_{d^*}) \leq \psi(C_d)$
for all Schur-convex functions ψ nonincreasing in their arguments and all designs d in Ω .

Letting $\psi(C_d)$ be $-\log \sum_{1=1}^{n} \mu_{d1}$, $\sum_{1=1}^{n} \mu_{d1}^{-1}$, $-\mu_{d1}$,

we obtain the well-known functions associated with D-, A- and E-optimality criteria, respectively. The above functions are Schur-convex because they all are symmetric and convex.

$\psi(C_d) = \sum\limits_{i=1}^{n} f(u_{di})$ with f convex nonincreasing and the ϕ_p criteria defined by Kiefer (1974) are also Schur-convex for the same reason. It is clear that all these functions are nonincreasing in their arguments. Note that the E-criterion is not a limiting case when formulated in terms of Schur-convexity.

Schur-optimality is a very strong criterion, as it implies all of the above. As the above examples indicate, it it probably quite satisfactory to look at just symmetric and convex functions of eigenvalues, rather than Schur-convex. But the techniques that we shall outline readily apply to Schur-convex functions and this motivates the extension.

3. ON AVERAGING AND MAJORIZATION

The principal tool that we shall employ when searching for Schur-optimal designs is contained in Theorem 3.2. This theorem relies, in turn, on a fundamental result on majorization due to Hardy, Littlewood and Polya (1934), which was later extended by Ostrowski (1952).

Whenever we write $x \preccurlyeq y$ for two vectors $x=(x_1,x_2,\ldots,x_n)$ and $y=(y_1,y_2,\ldots,y_n)$ in I^n we assume that

$$x_1 \leq x_2 \leq \cdots \leq x_n \ , \ y_1 \leq y_2 \leq \cdots \leq y_n$$

and that $\sum\limits_{i=1}^{m} y_i \leq \sum\limits_{i=1}^{m} x_i$ for all $1 \leq m \leq n$.

If $x \preccurlyeq y$ we say that y majorizes x.

The useful concept of majorization has been considered in the context of design optimality by Cheng (1979) to block designs with 4 varieties and an arbitrary number of blocks.

The following result can be easily derived from Ostrowski (1952):

<u>Theorem 3.1.</u> Let $x, y \in I^n$. If $x \preccurlyeq y$ then $\psi(x) \leq \psi(y)$ for all Schur-convex functions ψ nonincreasing in their arguments.

Denote by $\sigma(A)$ the nondecreasingly ordered vector of eigenvalues of the matrix A. The result we state next is essentially due to Ky Fan (1951) although he has not formulated it in this form.

<u>Lemma 3.1.</u> Let A_i $(1 \leq i \leq m)$ be nonnegative definite matrices. Then

$$\sigma\left(\sum_i \alpha_i A_i \right) \preccurlyeq \sum_i \alpha_i \, \sigma(A_i)$$

where $0 \leq \alpha_i \leq 1$ and $\sum \alpha_i = 1$.

Before we proceed, we need to introduce some notation. For a $v \times v$ matrix A and a permutation σ on the symbols $1, 2, \ldots, v$ we denote by A^σ the matrix obtained from A after performing the row and (same) column permutations as indicated by σ. That is $A^\sigma = PAP'$, where P is the $v \times v$ matrix representation of σ. Since A and A^σ are

similar matrices, their ordered spectra $\sigma(A)$ and $\sigma(A^\sigma)$ are identical for any permutation σ . We say that A^σ is a conjugate of A.

The following is an immediate but useful consequence (see also Magda (1979), Lemma 3.2.1).

Propostion 3.1. Any convex combination $\Sigma \, \alpha_i A^{\sigma_i}$ of conjugates of a nonnegative definite matrix A satisfies

$$\sigma(\sum_i \alpha_i A^{\sigma_i}) \preccurlyeq \sigma(A)$$

Now we can state and prove

Theorem 3.2. A design $d^* \in \Omega$ is Schur-better than d ($d^* \succcurlyeq d$) if

$$\sigma(C_{d*}) \preccurlyeq \sigma(\sum_i \alpha_i C_d^{\sigma_i})$$

ior some convex combination of conjugates of C_d.

Proof. Firstly, observe that by Theorem 3.1 $\mu(C_{d*}) \preccurlyeq \mu(C_d)$ implies $d^* \succcurlyeq d$. Since the last n components of $\mu(C_d)$ and $\sigma(C_d)$ are the same and the first $v-n$ components of $\sigma(C_d)$ are zero (for all $d \in \Omega$) $\mu(C_{d*}) \preccurlyeq \mu(C_d)$ is equivalent with $\sigma(C_{d*}) \preccurlyeq \sigma(C_d)$. So $\sigma(C_{d*}) \preccurlyeq \sigma(C_d)$ implies $d^* \succcurlyeq d$. Using the assumption and then Proposition 3.1 we have

$$\sigma(C_{d*}) \preccurlyeq \sigma(\sum_i \alpha_i C_d^{\sigma_i}) \preccurlyeq \sum_i \alpha_i \, \sigma(C_d^{\sigma_i}) = \sigma(C_d).$$

The last equality is true because $C_d^{\sigma_i}$ and C_d are conjugates and hence have the same spectrum. This concludes the proof.

Theorem 3.2 is helpful in the following way. Suppose
that a design $d* \in \Omega$ with a lot of symmetries is believed
optimal in some broad sense for the intuitive reasons men-
tioned in the introduction. It would be very satisfactory
to show that for a large class of designs $d \in \Omega$, $d*$ is
Schur-better than d (i.e. $d* \succcurlyeq d$). Because of its balance
$d*$ has an information matrix for which $\sigma(C_{d*})$ can be
computed. But $\sigma(C_d)$ for an arbitrary d is impossible
to calculate. It is very often possible, however, to compute

$\sigma(\sum_i \alpha_i \sigma(C_d^{\sigma_i}))$ for a convex combination of certain conju-
gates of C_d. The entries in the convex combination tend
to even out and this generally facilitates the computation of
the spectrum. The spectrum of the convex combination is a
helpful intermediary between $\sigma(C_{d*})$ and $\sigma(C_d)$ and the
content of Theorem 3.2 becomes of assistance. We shall il-
lustrate this in the next section.

Of particular importance is the average $\bar{C}_d = \frac{1}{m} \sum_{i=1}^{m} C_d^{\sigma_i}$

which we call an averaged version of C_d. We will be making
extensive use of averaged versions in the next section and
find it convenient to rely on the following:

<u>Theorem 3.3.</u> A design $d* \in \Omega$ is Schur-better than d if

$\sigma(C_{d*}) \leq \sigma(\overline{C}_d)$ for some averaged version $\overline{C}_d$ of C_d.

Clearly $d*$ is Schur-optimal over Ω if it is Schur-better than all the designs d in Ω.

4. RESULTS ON SCHUR-OPTIMALITY

Let $d* \in \Omega$ be a design for which $\mu(C_{d*})$ has all its entries equal to μ_{d*}. Assume also that the trace of C_d is at most equal to the trace of C_{d*}, for all $d \in \Omega$. We claim that $\mu(C_{d*}) \preccurlyeq \mu(C_d)$. Let $\mu(C_d) = (\mu_{d1}, \ldots, \mu_{dn})'$.

Then if $\sum\limits_{i=1}^{n} \mu_{di} > k\mu_{d*}$ for some $k \leq n$ we have $\mu_{di} \geq \mu_{d*}$

for all $i \geq k + 1$ and hence also $\sum\limits_{i=k+1}^{n} \mu_{di} \geq (n-k)\mu_{d*}$.

This implies that $\text{tr}C_d = \sum\limits_{i=1}^{k} \upsilon_{di} + \sum\limits_{i=k+1}^{n} \upsilon_{di} > k\mu_{d*} + (n-k)\mu_{d*} = n\mu_{d*} = \text{tr}C_{d*}$, a contradiction. We therefore have $\mu(C_{d*}) \preccurlyeq \mu(C_d)$. Theorem 3.1 gives now

Theorem 4.1. If there exists a design $d*$ in Ω such that $\mu(C_{d*})$ has all its entries equal and trace $C_{d*} \geq$ trace C_d, for all $d \in \Omega$, then $d*$ is Schur-optimal over Ω.

There are two particularly useful consequences. Before we state them, let us call a matrix completely symmetric if all its diagonal entries are equal and all its off-diagonals are also equal. The following two propositions readily

satisfy the assumptions of Theorem 4.1.

Propostion 4.1. Let $\cap$ consist of designs d for which C_d has zero row sums. Then a design d* ϵ $\cap$ with a completely symmetric matrix of maximal trace is Schur-optimal over $\cap$.

Propostion 4.2. If d* ϵ $\cap$ is a design whose information matrix is a multiple of the identity matrix and has maximal trace, then d* is Schur-optimal over $\cap$.

The above propostions are reformulations of Proposition 1 and Proposition 1' of Kiefer (1975) in terms of Schur-optimality. Kiefer phrased the aforementioned results in terms of Universal-optimality, a concept that proves to be very valuable especially when dealing with optimality in regular settings which permit symmetric designs. The essential difference between Universal-optimality and Schur-optimality lies in the relaxation of monotonicity in a scalar to that in each individual component of $\mu(C_d)$. This permits immediate connections to the results on hermitian matrices by Ky Fan and results on majorization by Hardy, Littlewood and Polya. Schur-optimality is applicable in less regular settings especially when showing that a design with desirable symmetries is Schur-better than large classes of less symmetric designs. We illustrate this next.

Throughout the remainder of the paper, Ω is assumed to consist of designs d for which the information matrix C_d has row sums zero. Moreover, $u(C_d)$ will be a vector with $v-1$ components, as is the case in most discrete settings.

For a $m \times m$ matrix $C = (c_{ij})$, let us denote by $\Delta(C)$ the quantity $(m-1) \sum_{i=1}^{m} c_{ii} - \sum_{i \neq j} c_{ij}$. We can now state and prove

Theorem 4.2. If Ω contains a design $d*$ such that $u(C_{d*})$ has exactly two distinct components, $\mu_{d*1} < \mu_{d*2}$, with μ_{d*1} of multiplicity r and μ_{d*2} of multiplicity s ($r+s = v-1$), then $d*$ is Schur-optimal over the collection of designs d in Ω which satisfy Trace $C_d \leq$ Trace C_{d*} and either (a) or (b) below:

(a) $\Delta(M_d) \leq r(r+1) \mu_{d*1}$ for some $(r+1) \times (r+1)$ principal minor M_d of C_d.

(b) $\Delta(M_d) \geq s(s+1) \mu_{d*2}$ for some $(s+1) \times (s+1)$ principal minor M_d of C_d.

Proof: Let $d \in \Omega$ satisfy (a). Write the information matrix C_d in such a way that M_d is in the upper left hand corner. Average C_d over the first $r+1$ rows and columns and then over the remaining $v-r-1$ rows and columns. Let

$\overline{C}_d$ be the average version of C_d so obtained. Explicitly

$$\overline{C}_d = \frac{1}{(r+1)!(v-r-1)!} \sum_i C_d^{\sigma_i}$$

where σ_i is a product of a permutation on the first r+1 rows and columns of C_i with a permutation on the last v-r-1 rows and columns. The shape of $\overline{C}_d$ is

$$\overline{C}_d = \begin{bmatrix} (a_d+\alpha_d)I - \alpha_d J & -\beta_d J \\ \\ -\beta_d J & (b_d+\gamma_d)I - \gamma_d J \end{bmatrix}$$

where the upper left hand corner is (r+1)×(r+1) and I and J are respectively the identity matrix and the matrix with all its entries 1. By adding the v×v matrix $\beta_d J$ to $\overline{C}_d$, the eigenvalues of $\overline{C}_d$ are found to be 0, $a_d+\alpha_d$, $v\overline{\beta}_d$ and $b_d+\gamma_d$ of multiplicities 1, r, 1 and v-r-2, respectively. By Propostion 3.1 all these eigenvalues are nonnegative. Next we show that $u(C_{d*}) \preccurlyeq u(\overline{C}_d)$. To achieve this let us denote by $\overline{u(\overline{C}_d)}$ the vector whose first r entries are equal to the average of the first r entries of $u(\overline{C}_d)$, and whose last s entries are equal to the average of the last s entries of $u(\overline{C}_d)$. It is easy to see that $\overline{u(\overline{C}_d)} \preccurlyeq \mu(\overline{C}_d)$. Since $\Delta(M_d) = r(r+1)(a_d+\alpha_d)$ and since d

satisfies (a) we have $a_d + \alpha_d \leq \underline{\mu_{d*1}}$. This implies that
each of the first r entries of $\mu(\overline{C}_d)$ is at most μ_{d*1}
Since Trace $C_d \leq$ Trace C_{d*}, it now follows without much
difficulty that $\mu(C_{d*}) \preceq \mu(\overline{C}_d)$. This shows that $\mu(C_{d*}) \preceq \mu(\overline{C}_d)$
and hence also that $\sigma(C_{d*}) \preceq \sigma(\overline{C}_d)$. By Theorem 3.3 we
now conclude that $d*$ is Schur-better than d. When $d \in \Omega$
satisfies (b) a similar averaging yields $d* \succeq d$. This con-
cludes the proof of the theorem.

One important remark is in order here. Theorem 4.2 does
not impose any condition on having μ_{d*1} and μ_{d*2} close
together. It is easy to see, however, that when they are close
to each other more designs satisfy (a) or (b) in the
theorem and hence $d*$ is Schur-optimal over a larger col-
lection of designs. A convenient way to ensure the closeness
of μ_{d*1} and μ_{d*2} is to demand that Trace$(C^2_{d*}) \leq$ Trace(C^2_d)
for all $d \in \Omega$. Whenever such a design exists, Theorem 4.2
ensures its Schur-optimality over large subfamilies of de-
signs in Ω . This is a helpful fact, as it eases the search
for the D-,A- and E-optimal designs in Ω .

Another important observation relates to the selection of
a principal minor M_d for which (b) holds. Since C_d
is nonnegative definite the diagonal elements of C_d are
relatively large. Moreover, the diagonal elements of M_d
carry a lot of weight (each one is multiplied by s) in $\Delta(M_d)$.

The off-diagonals of M_d are therefore of much less importance when it comes to maximizing $\Delta(M_d)$ over M_d . Whence it makes sense to first choose the principal minor M_d of <u>maximal</u> <u>trace</u> and check if condition (b) in the above theorem is satisfied for this particular principal minor.

There is one special case of Theorem 4.2 that deserves mention. It is the setting in which all the designs in Ω have information matrices with nonpositive off-diagonals. Such is the case in the setting of block designs or two way elimination, for example.

Denote by c_{dij} the entries of the information matrix C_d in such a way that $c_{d11} \leq c_{d22} \leq \ldots \leq c_{dv,v}$.

<u>Theorem 4.3.</u> Let Ω consist of designs whose information matrices have zero row sums and nonpositive off-diagonals. If Ω contains a design $d*$ which satisfies the assumptions of Theorem 4.2 then $d*$ is Schur-optimal over the class of designs d in Ω which satisfy either

$$\sum_{i=1}^{r+1} c_{dii} \leq r\, \mu_{d*1} \qquad \text{or} \qquad \sum_{i=v-s}^{v} c_{dii} \geq (s+1)\, \mu_{d*2}$$

<u>Proof:</u> Since $c_{dij} \leq 0$, for $i \neq j$, and the row sums of C_d are zero, we have $\Delta(M_d) = r \sum_{i=1}^{r+1} c_{dii} - \sum_{i \neq j} c_{dij} \leq$

$$\leq r \sum_{i=1}^{r+1} c_{dii} + \sum_{i=1}^{r+1} c_{dii} = (r+1) \sum_{i=1}^{r+1} c_{dii} \leq r(r+1)\, \mu_{d*1} \quad \text{where}$$

M_d is the $(r+1)\times(r+1)$ principal minor of C_d whose diagonal entries are c_{dii}, $i = 1,2,\ldots,r+1$. We thus satisfy condition (a) in Theorem 4.2, and hence $d^* \geq d$. By letting M_d be the $(s+1)\times(s+1)$ principal minor in the lower right hand corner of C_d we have (by simply using the fact that $c_{dij} \leq 0$ for $i \neq j$) $\Delta(M_d) \geq \sum_{i=v-s}^{v} c_{dii} \geq s(s+1)\mu_{d^*2}$. We are now done by (b) of Theorem 4.2. This ends the proof.

Theorem 4.2 and 4.3 have a number of consequences when considered under specific linear models. We shall examine next a corollary in the block design setting.

Order the replication numbers in a design $d \in \Omega_{v,b,k}$ such that $r_{d1} \leq r_{d2} \leq \cdots \leq r_{dv}$. By observing that the ith diagonal entry of the information matrix of a binary design design d is $\frac{k-1}{k}r_{di}$, one has the following reformulation of Theorem 4.2 in terms of the replication numbers r_{di}.

<u>Corollary 4.1.</u> If d^* is a binary design in $\Omega_{v,b,k}$ whose information matrix has exactly two distinct nonzero eigenvalues, $\mu_{d^*1} < \mu_{d^*2}$ the former being of multiplicity r and the latter of multiplicity s $(r+s = v-1)$, then d^* is Schur-better than all the binary designs d which satisfy either

$$\sum_{i=1}^{r+1} r_{di} \leq \frac{kr}{k-1}\mu_{d^*1} \quad \text{or} \quad \sum_{i=v-s}^{v} r_{di} \geq \frac{k(s+1)}{k-1}\mu_{d^*2}.$$

Designs d* which satisfy the assumptions of
Theorem 4.1 exist in many settings. Among block designs
we mention the Partially Balanced Incomplete Block designs
with two associate classes, extended and abridged Balanced
Incomplete Block designs with any number of disjoint bi-
nary blocks and Group Divisible designs with $\lambda_2 = \lambda_1 \pm 1$
adjointed or abridged by disjoint binary blocks compatible
with the partition of the groups (see Constantine (1980)).
When a pair of varieties occurs in either λ or $\lambda+1$
blocks (which ensures the closeness of μ_{d*1} and μ_{d*2}),
Theorem 4.2 asserts that all these designs are Schur-optimal
over large classes of designs. We now examine one such
instance more closely.

Let $d* \in \Omega_{v,b,k}$ be a Group Divisible design with $\lambda_2 = \lambda_1 + 1$
(where λ_1 is the number of blocks containing two varieties
that are in the same group) and m groups of size n.
Examples show that d* is not Schur-optimal (but very likely
both D- and A-optimal) over all designs. For m=2 a
very strong optimality statement (including D- and A-)
was proved by Cheng (1978); the E-optimality has been ob-
tained by Takeuchi (1961) for general m.

For a design $d \in \Omega_{v,b,k}$ let $N_d N_d' = (\lambda_{dij})$. Let
furthermore $(M_i)_{i=1,\ldots,m}$ be an arbitrary partition of

the v varieties in m groups of size n each and set

$$r_d = \frac{1}{v} \sum_{i=1}^{v} (k\, r_{di} - \lambda_{dii}) \quad \text{and} \quad \lambda_d = \frac{1}{v(n-1)} \sum_{t=1}^{m} \sum_{\substack{i \neq j \\ i,j \in M_t}} \lambda_{dij}.$$

It is not very hard to show that the matrix $k\overline{C}_d$ having
m diagonal blocks of size n, all equal to $(r_d + \lambda_d)I - \lambda_d J$
and with all the entries outside these blocks equal, is an
averaged version of kC_d. One can show (see Magda(1979))
that $r_d + \lambda_d$ is an eigenvalue of $k\overline{C}_d$ and has multiplicity
v-m. Since d* is binary, we have $r_d \leq (k-1)r$ where r
is the replication of any variety in d*. Now, if $\lambda_d \leq \lambda_1$
we have $r_d + \lambda_d \leq (k-1)r + \lambda_1$. But $(k-1)r + \lambda_1$
is the smallest eigenvalue of kC_{d*} and it has multiplicity
v-m. With the assumption that $\lambda_d \leq \lambda_1$ and using the fact
that kC_{d*} has maximal trace, it can be now readily ver-
ified that $\sigma(C_{d*}) \preccurlyeq \sigma(\overline{C}_d)$. By Theorem 3.3 we therefore
have:

<u>Theorem 4.3.</u> A Group Divisible design $d* \in \Omega_{v,b,k}$ with
m groups of size n and $\lambda_2 = \lambda_1 + 1$ is Schur-optimal
over the class of designs d which satisfy $\lambda_d \leq \lambda_1$
for λ_d associated with some partition of the varieties.

REFERENCES

Cheng, C.S. (1978). Optimality of certain asymmetrical experimental designs. Ann. Statist., 6, 1239-1261.

Cheng, C.S. (1979). Optimal incomplete block designs with four varieties. To appear in Sankhya.

Constantine, G.M. (1979). Some E-optimal block designs. Submitted to the Annals of Statistics.

Ehrenfeld, S. (1955). On the efficiencies of experimental designs. Ann. Math. Statist., 26, 247-255.

Fan, K. (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. Proc. Nat. Acad. Sci. U.S.A., 37, 760-766.

Hardy, G.H., Littlewood, J.E., and Polya, G. (1934). Inequalities. Cambridge University Press, London and New York.

Kiefer, J. (1959). Optimum experimental designs. J. Roy. Stat. Soc. Ser. B., 21, 272-304.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). Ann. Math. Statist., 5, 849-879.

Kiefer, J. (1975). Construction and optimality of generalized Youden designs. In: A Survey of Statistical Designs and Linear Models (J.N. Srivastava, ed.), 333-353, North-Holland, Amsterdam.

Magda, C.G. (1979). On E-optimal block designs and Schur-optimality. Thesis. University of Illinois at Chicago Circle.

Ostrowski, A.M. (1951). Sur quelques applications des fonctions convexes et concaves an sens de I. Schur. J. Math. Pures Appl., 9 , 31, 253-292.

Takeuchi, K. (1961). On the optimality of certain types of PBIB designs. Rep. Statist. Appl. Res. JUSE., 8, 140-145.

END

DATE
FILMED

8/

DTIC